

SID1

not prior

RESULT 5  
AL365193/c  
LOCUS AL365193 51887 bp DNA linear PRI 22-NOV-2000  
DEFINITION Human DNA sequence from clone RP3-526N11 on chromosome 6, complete  
sequence.  
ACCESSION AL365193  
VERSION AL365193.12 GI:11322849  
KEYWORDS HTG. earliest 7/8/00 =  
SOURCE human. not prior  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 51887)  
AUTHORS Tracey, A.  
TITLE Direct Submission  
JOURNAL Submitted (22-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk  
COMMENT On Nov 23, 2000 this sequence version replaced gi:11190578.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
on the WORMPEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr6>  
RP3-526N11 is from the library RPCI-3 constructed at the Roswell  
Park Cancer Institute by the group of Pieter de Jong. For further  
details see <http://bacpac.med.buffalo.edu/>  
VECTOR: pCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone  
RP3-526N11 It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone RP3-466P17 is at 51785 in this sequence.  
The true right end of clone RP1-28C20 is at 100 in this sequence.  
FEATURES  
source Location/Qualifiers  
1. .51887  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP3-526N11"  
/clone\_lib="RPCI-3"  
repeat\_region 2. .1199

misc_feature	/note="L1M4 repeat: matches 2626. .3805 of consensus" complement(1106. .1664)
	/note="match: GSS: Em:AQ506127"
repeat_region	2390. .2538
	/note="L1 repeat: matches 4015. .4167 of consensus"
repeat_region	3459. .3774
	/note="AluYb8 repeat: matches 1. .316 of consensus"
repeat_region	3948. .4156
	/note="HAL1 repeat: matches 511. .747 of consensus"
repeat_region	4168. .5043
	/note="L1PA14 repeat: matches 5266. .6149 of consensus"
repeat_region	5077. .5807
	/note="L1PA14 repeat: matches 4535. .5263 of consensus"
repeat_region	5804. .6152
	/note="HAL1 repeat: matches 170. .526 of consensus"
repeat_region	7143. .7294
	/note="LTR41 repeat: matches 1. .175 of consensus"
repeat_region	7866. .7927
	/note="31 copies 2 mer tt 69% conserved"
repeat_region	8400. .8570
	/note="MER69B repeat: matches 1047. .1225 of consensus"
repeat_region	8671. .9549
	/note="MER69B repeat: matches 1. .914 of consensus"
misc_feature	9600. .10107
	/note="match: GSS: Em:AQ459097"
repeat_region	9960. .10172
	/note="L2 repeat: matches 2522. .2749 of consensus"
repeat_region	13463. .13792
	/note="MER93 repeat: matches 67. .397 of consensus"
repeat_region	14604. .14914
	/note="MER44A repeat: matches 51. .333 of consensus"
repeat_region	16516. .16811
	/note="AluSg1 repeat: matches 1. .296 of consensus"
repeat_region	19348. .21583
	/note="TIGGER1 repeat: matches 1. .2249 of consensus"
repeat_region	21704. .22053
	/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region	22079. .23644
	/note="THE1B-INTERNAL repeat: matches 5. .1580 of consensus"
repeat_region	23645. .23993
	/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region	24238. .24531
	/note="AluJb repeat: matches 1. .290 of consensus"
repeat_region	24731. .24990
	/note="L2 repeat: matches 2229. .2500 of consensus"
repeat_region	24991. .25362
	/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region	25363. .25552
	/note="L2 repeat: matches 2044. .2229 of consensus"
misc_feature	25534. .25976
	/note="match: GSS: Em:AQ786188"
repeat_region	26695. .27052
	/note="MER1B repeat: matches 1. .337 of consensus"
repeat_region	27392. .28695
	/note="L1MA9 repeat: matches 4980. .6308 of consensus"
repeat_region	28981. .29439

```

repeat_region    /note="MLT1C repeat: matches 1. .459 of consensus"
                 29528. .29685
repeat_region    /note="MIR repeat: matches 93. .262 of consensus"
                 29702. .30591
                 /note="L2 repeat: matches 1798. .2716 of consensus"
misc_feature     31958. .32309
                 /note="match: STS: Em:G24207"
repeat_region    32486. .32705
                 /note="MIR repeat: matches 10. .248 of consensus"
repeat_region    33224. .33292
                 /note="MIR repeat: matches 57. .126 of consensus"
misc_feature     complement(34798. .35362)
                 /note="match: GSS: Em:AQ627350"
misc_feature     complement(34924. .35379)
                 /note="match: GSS: Em:AQ127321"
repeat_region    36912. .37203
                 /note="AluSx repeat: matches 1. .290 of consensus"
repeat_region    37323. .37548
                 /note="MIR repeat: matches 6. .252 of consensus"
repeat_region    38453. .38482
                 /note="15 copies 2 mer ca 100% conserved"
repeat_region    40441. .40644
                 /note="MIR repeat: matches 22. .235 of consensus"
repeat_region    41172. .41635
                 /note="L2 repeat: matches 2264. .2709 of consensus"
misc_feature     41939. .42389
                 /note="match: GSS: Em:AQ126602"
repeat_region    42301. .42460
                 /note="L1M4 repeat: matches 3751. .3918 of consensus"
repeat_region    42800. .43080
                 /note="AluSx repeat: matches 1. .295 of consensus"
repeat_region    43229. .43533
                 /note="AluSx repeat: matches 6. .309 of consensus"
repeat_region    43546. .43712
                 /note="MIR repeat: matches 71. .240 of consensus"
repeat_region    45906. .46019
                 /note="MER5B repeat: matches 66. .178 of consensus"
misc_feature     complement(47728. .48255)
                 /note="match: GSS: Em:AQ786765"
repeat_region    47743. .48126
                 /note="MLT1B repeat: matches 1. .388 of consensus"
misc_feature     48248. .48771
                 /note="match: GSS: Em:AQ680415"
misc_feature     48263. .48754
                 /note="match: GSS: Em:AQ205491"
repeat_region    49005. .49762
                 /note="L2 repeat: matches 1890. .2742 of consensus"
misc_feature     50377. .50863
                 /note="match: GSS: Em:AQ478674"
misc_feature     complement(50429. .50848)
                 /note="match: GSS: Em:AQ525841"
repeat_region    50541. .50635
                 /note="MER5A repeat: matches 4. .96 of consensus"
BASE COUNT      16111 a   9944 c   9493 g  16339 t
ORIGIN

```

Query Match

73.4%; Score 2297.4; DB 9; Length 51887;

Best Local Similarity 99.4%; Pred. No. 0;  
Matches 2390; Conservative 0; Mismatches 6; Indels 9; Gaps 8;

SID 1 102(b) against hybridizing.

RESULT 12

AAT18566

ID AAT18566 standard; cDNA; 181 BP.

XX

AC AAT18566;

XX

DT 05-NOV-1996 (first entry)

XX

DE Human chondrocyte RNA derived cDNA mol. TAU 10(1).

XX

KW Human; chondrocyte; gene specific; primer; probe; isolation;

KW interleukin-1beta; IL-1beta; diagnosis; connective tissue disease;

KW osseteoarthritis; rheumatoid arthritis; ss.

XX

OS Homo sapiens.

XX

PN EP705842-A2.

XX

PD 10-APR-1996.

XX

PF 02-OCT-1995; 95EP-0115510.

XX

PR 06-OCT-1994; 94EP-0115751.

XX

PA (FARH ) HOECHST AG.

XX

PI Bartnik E, Margerie D;

XX

DR WPI; 1996-181045/19.

XX

PT Diagnosis and treatment of IL-1 mediated connective tissue diseases  
PT - using osteopontin, calnexin, TSG-6 gene prod., genes encoding them  
PT or antibodies to them

XX

PS Claim 10; Page 25; 31pp; English.

XX

CC The present sequence is 1 of 52 cDNA mols. derived from human  
CC chondrocyte RNA, which were searched against DNA databases for  
CC homology to known human genes; in this case no homology was found.  
CC The DNA mols. can be used for the prodn. of gene specific primers  
CC and probes to isolate genes induced by treating (esp. human)  
CC chondrocytes with interleukin-1beta (IL-1beta), and for the  
CC diagnosis of IL-1beta related connective tissue diseases, in  
CC partic. osseteoarthritis or rheumatoid arthritis.

XX

SQ Sequence 181 BP; 63 A; 28 C; 35 G; 55 T; 0 other;

Query Match 5.1%; Score 158; DB 17; Length 181;

Best Local Similarity 98.9%; Pred. No. 4.4e-25;

Matches 180; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 2669 GGAGATGACATTTGCTTTGGGCAGAGGCAGCTAGCCAGGACACATTTCCACTATAATTTT 2728

Db 1 GGAGATGACATTTGCTTTGGGCAGAGGCAGCTAGCCAGGACACATTTCCACTATAATTTT 60

Qy 2729 ACAAAGTTAAATTTATAAGCTAGCATTAAGTAAAGTGAAG-TCCAGCTCCCTTGCTAAAA 2787

|||||  
Db 61 ACAAAGTTAAATTTATAAGCTAGCATTAAAGTAAAGTGAAGTTCCAGCTCCCTTGCTAAAA 120  
Qy 2788 ATAAGTAGAGGTAATAATTGGTATTCAGGTAAC TCATTACAGTCATAATGTGTTGTGAA 2847  
|||||  
Db 121 ATAAGTAGAGGTAATAATTGGTATTCAGGTAAC TCATTACA-TCATAATGTGTTGTGAA 179  
Qy 2848 AA 2849  
||  
Db 180 AA 181

SID 3

Query Match 5.4%; Score 158; DB 17; Length 181;  
Best Local Similarity 98.9%; Pred. No. 3.4e-31;  
Matches 180; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```
Qy 2481 GGAGATGACATTTGCTTTGGGCAGAGGCAGCTAGCCAGGACACATTTCCACTATAATTTT 2540
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGAGATGACATTTGCTTTGGGCAGAGGCAGCTAGCCAGGACACATTTCCACTATAATTTT 60

Qy 2541 ACAAAGTTAAATTTATAAGCTAGCATTAAAGTAAAGTGAAG-TCCAGCTCCCTTGCTAAAA 2599
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ACAAAGTTAAATTTATAAGCTAGCATTAAAGTAAAGTGAAGTTCCAGCTCCCTTGCTAAAA 120

Qy 2600 ATAAGTAGAGGTAATAATTGGTATTCAGGTAACTCATTTACAGTCATAATGTGTTGTGAA 2659
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATAAGTAGAGGTAATAATTGGTATTCAGGTAACTCATTTACA-TCATAATGTGTTGTGAA 179

Qy 2660 AA 2661
      ||
Db   180 AA 181
```

Query Match 14.9%; Score 136.6; DB 17; Length 181;  
Best Local Similarity 96.4%; Pred. No. 1.8e-22;  
Matches 161; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

[illegible]



Seq ID 1

not prior art

RESULT 1  
 AF084535 3128 bp mRNA linear PRI 01-OCT-1999  
 LOCUS  
 DEFINITION Homo sapiens laforin (EPM2A) mRNA, complete cds.  
 ACCESSION AF084535  
 VERSION AF084535.2 GI:6005985  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 3128)  
 AUTHORS Minassian, B.A., Lee, J.R., Herbrick, J.A., Huizenga, J., Soder, S.,  
 Mungall, A.J., Dunham, I., Gardner, R., Fong, C.Y., Carpenter, S.,  
 Jardim, L., Satishchandra, P., Andermann, E., Snead, O.C. III,  
 Lopes-Cendes, I., Tsui, L.C., Delgado-Escueta, A.V., Rouleau, G.A. and  
 Scherer, S.W.  
 TITLE Mutations in a gene encoding a novel protein tyrosine phosphatase  
 cause progressive myoclonus epilepsy  
 JOURNAL Nat. Genet. 20 (2), 171-174 (1998)  
 MEDLINE 98442653  
 PUBMED 9771710  
 REFERENCE 2 (bases 1 to 3128)  
 AUTHORS Lee, J.R. and Scherer, S.W.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-AUG-1998) Department of Genetics, The Hospital for  
 Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada  
 REFERENCE 3 (bases 1 to 3128)  
 AUTHORS Minassian, B.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-OCT-1999) Department of Genetics, The Hospital for  
 Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada  
 REMARK Sequence update by submitter  
 COMMENT On Oct 1, 1999 this sequence version replaced gi:3978461.  
 FEATURES  
 source Location/Qualifiers  
 1. .3128  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 gene 1. .3128  
 /gene="EPM2A"  
 CDS 1. .996  
 /gene="EPM2A"  
 /codon\_start=1  
 /product="laforin"  
 /protein\_id="AAC83347.2"  
 /db\_xref="GI:6005986"  
 /translation="MRFRFGVVVPPAVAGARPELLVVGSRPELGRWEPRGAVRLRPAG  
 TAAGDGALALQEPGLWLGEVELAAEEAAQDGAEPGRVDTFWYKFLKREPGGELSWEGN  
 GPHHDCCTYNENNLVDGVYCLPIGHWIEATGHTNEMKHTTDFYFNIAGHQAMHYSRI  
 LPNIWLGSCPRQVEHVTIKLKHGELGITAVMNFQTEWDIVQNSSGCNRYPEPMTPTDTMI  
 KLYREEGLAYIWMPTPDMSTEGRVQMLPQAVCLLHALLKKGHIYVYVHCNAGVGRSTAA  
 VCGWLQYVMGWNLRKVQYFLMAKRPVYIDEEALARAQEDFFQKFGKVRSSVCSL"  
 3'UTR 997. .3128  
 /gene="EPM2A"  
 BASE COUNT 877 a 643 c 763 g 845 t  
 ORIGIN

Query Match 99.9%; Score 3126.4; DB 9; Length 3128;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 ATGCGCTTCCGCTTTGGGGTGGTGGTGCCACCCGCCGTGGCCGGCGCCCGGCCGGAGCTG 60
      |||
Db      1 ATGCGCTTCCGCTTTGGGGTGGTGGTGCCACCCGCCGTGGCCGGCGCCCGGCCGGAGCTG 60

Qy     61 CTGGTGGTGGGGTCGCGGCCCGAGCTGGGGCGTTGGGAGCCGCGCGGTGCCGTCCGCCTG 120
      |||
Db     61 CTGGTGGTGGGGTCGCGGCCCGAGCTGGGGCGTTGGGAGCCGCGCGGTGCCGTCCGCCTG 120

Qy    121 AGGCCGGCCGGCACCCGCGGCGGGCGACGGGGCCCTGGCGCTGCAGGAGCCGGGCCTGTGG 180
      |||
Db    121 AGGCCGGCCGGCACCCGCGGCGGGCGACGGGGCCCTGGCGCTGCAGGAGCCGGGCCTGTGG 180

Qy    181 CTCGGGGAGGTGGAGCTGGCGGCCGAGGAGCGGCGCAGGACGGGGCGGAGCCGGGCCGC 240
      |||
Db    181 CTCGGGGAGGTGGAGCTGGCGGCCGAGGAGCGGCGCAGGACGGGGCGGAGCCGGGCCGC 240

Qy    241 GTGGACACGTTCTGGTACAAGTTCCTGAAGCGGGAGCCGGGAGGAGAGCTCTCCTGGGAA 300
      |||
Db    241 GTGGACACGTTCTGGTACAAGTTCCTGAAGCGGGAGCCGGGAGGAGAGCTCTCCTGGGAA 300

Qy    301 GGCAATGGACCTCATCATGACCGTTGCTGTACTTACAATGAAAACAACCTTGGTGGATGGT 360
      |||
Db    301 GGCAATGGACCTCATCATGACCGTTGCTGTACTTACAATGAAAACAACCTTGGTGGATGGT 360

Qy    361 GTGTATTGTCTCCCAATAGGACACTGGATTGAGGCCACTGGGCACACCAATGAAATGAAG 420
      |||
Db    361 GTGTATTGTCTCCCAATAGGACACTGGATTGAGGCCACTGGGCACACCAATGAAATGAAG 420

Qy    421 CACACAACAGACTTCTATTTTAATATTGCAGGCCACCAAGCCATGCATTATTCAAGAATT 480
      |||
Db    421 CACACAACAGACTTCTATTTTAATATTGCAGGCCACCAAGCCATGCATTATTCAAGAATT 480

Qy    481 CTACCAAATATCTGGCTGGGTAGCTGCCCTCGTCAGGTGGAACATGTTACCATCAAACCTG 540
      |||
Db    481 CTACCAAATATCTGGCTGGGTAGCTGCCCTCGTCAGGTGGAACATGTAACCATCAAACCTG 540

Qy    541 AAGCATGAATTGGGGATTACAGCTGTAATGAATTTCCAGACTGAATGGGATATTGTACAG 600
      |||
Db    541 AAGCATGAATTGGGGATTACAGCTGTAATGAATTTCCAGACTGAATGGGATATTGTACAG 600

Qy    601 AATTCCTCAGGCTGTAACCGCTACCCAGAGCCCATGACTCCAGACACTATGATTAAACTA 660
      |||
Db    601 AATTCCTCAGGCTGTAACCGCTACCCAGAGCCCATGACTCCAGACACTATGATTAAACTA 660

Qy    661 TATAGGGAAGAAGGCTTGGCCTACATCTGGATGCCAACACCAGATATGAGCACCGAAGGC 720
      |||
Db    661 TATAGGGAAGAAGGCTTGGCCTACATCTGGATGCCAACACCAGATATGAGCACCGAAGGC 720

Qy    721 CGAGTACAGATGCTGCCCCAGGCGGTGTGCCTGCTGCATGCGCTGCTGGAGAAGGGACAC 780
      |||
Db    721 CGAGTACAGATGCTGCCCCAGGCGGTGTGCCTGCTGCATGCGCTGCTGGAGAAGGGACAC 780

Qy    781 ATCGTGACGTGCACTGCAACGCTGGGGTGGGCGGCTCCACCGGGCTGTCTGCGGCTGG 840
```

Db	781		ATCGTGTACGTGCAACGCTGGGGTGGGCCGCTCCACCGCGGCTGTCTGCGGCTGG	840
Qy	841		CTCCAGTATGTGATGGGCTGGAATCTGAGGAAGGTGCAGTATTTCTCATGGCCAAGAGG	900
Db	841		CTCCAGTATGTGATGGGCTGGAATCTGAGGAAGGTGCAGTATTTCTCATGGCCAAGAGG	900
Qy	901		CCGGCTGTCTACATTGACGAAGAGGCCTTGGCCCCGGGCACAAGAAGATTTTTTCCAGAAA	960
Db	901		CCGGCTGTCTACATTGACGAAGAGGCCTTGGCCCCGGGCACAAGAAGATTTTTTCCAGAAA	960
Qy	961		TTTGGGAAGGTTTCGTTCTTCTGTGTGTAGCCTGTAGCTGGTCAGCCTGCTTCTGCCCCCT	1020
Db	961		TTTGGGAAGGTTTCGTTCTTCTGTGTGTAGCCTGTAGCTGGTCAGCCTGCTTCTGCCCCCT	1020
Qy	1021		CCTGATTTCCCTAAGGAGCCTGGGATGATGTTGGTCAAATGACCTAGAAACAAGGATTCT	1080
Db	1021		CCTGATTTCCCTAAGGAGCCTGGGATGATGTTGGTCAAATGACCTAGAAACAAGGATTCT	1080
Qy	1081		ACCTGAACTGAAAGGACTGTGTGACCTCCCCAAGCCAACCACTTTCACCTGGGATGACTT	1140
Db	1081		ACCTGAACTGAAAGGACTGTGTGACCTCCCCAAGCCAACCACTTTCACCTGGGATGACTT	1140
Qy	1141		TCGATTATGCTTTGGTTTGGGGCTGTATTTTTGAAATACTCTACAAGAAAGCTGTGGCTC	1200
Db	1141		TCGATTATGCTTTGGTTTGGGGCTGTATTTTTGAAATACTCTACAAGAAAGCTGTGGCTC	1200
Qy	1201		AACACATGAGAAGAAGCACGAAGCAGTTAGGCTGTACATCAGACAGAAGGGTAATGCGTG	1260
Db	1201		AACACATGAGAAGAAGCACGAAGCAGTTAGGCTGTACATCAGACAGAAGGGTAATGCGTG	1260
Qy	1261		CAGTTCCTGCTGCCTGCAGGCAGACGAGGCCTTTGCTTTACAGCACTGTATGTGTTGCAC	1320
Db	1261		CAGTTCCTGCTGCCTGCAGGCAGACGAGGCCTTTGCTTTACAGCACTGTATGTGTTGCAC	1320
Qy	1321		GATGGATCCGTGACAGCACTTTCCTGTTGCACTGAAACTCTTGGCCATGTAGAGGAAAAG	1380
Db	1321		GATGGATCCGTGACAGCACTTTCCTGTTGCACTGAAACTCTTGGCCATGTAGAGGAAAAG	1380
Qy	1381		ATATGGAGTTATGTGGATTTTCATCACTAGTATGTGTGCCGTGAGCTGGTCAGTTGCCAAA	1440
Db	1381		ATATGGAGTTATGTGGATTTTCATCACTAGTATGTGTGCCGTGAGCTGGTCAGTTGCCAAA	1440
Qy	1441		GGAGGAAATAAGGTTAGAAGCCTGAACCGTTACAAAAGAAGAGCTCACTATGGTCAAAAA	1500
Db	1441		GGAGGAAATAAGGTTAGAAGCCTGAACCGTTACAAAAGAAGAGCTCACTATGGTCAAAAA	1500
Qy	1501		GTGATGGCTTTTCAGGACTTGTTTTTTATCCTGCCTCACAGTTGTTAAAGTCTGTTCCAAG	1560
Db	1501		GTGATGGCTTTTCAGGACTTGTTTTTTATCCTGCCTCACAGTTGTTAAAGTCTGTTCCAAG	1560
Qy	1561		GCATCACCTTCCTTCTCTACCCAACAACCCTGTGTAACAATAAAGTAGAATTATCTCTC	1620
Db	1561		GCATCACCTTCCTTCTCTACCCAACAACCCTGTGTAACAATAAAGTAGAATTATCTCTC	1620
Qy	1621		ATTTGTTGGTGGTTTTTCTCAAATACCAAACAAAGCAAAAAATACCCTTGTTTTTTTA	1680

Db 1621 ATTTGTTGGTGGTTTTTCCTCAAAATTACCAAACAAAGCAAAAAATACCCTTGTTTTTTA 1680  
 Qy 1681 TAGTTGAGATGTCAAGGAAGTTAAATTGAGGCTTAATGAGCATAGGTAGCTTGTTCCAAGG 1740  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1681 TAGTTGAGATGTCAAGGAAGTTAAATTGAGGCTTAATGAGCATAGGTAGCTTGTTCCAAGG 1740  
 Qy 1741 TCTCATGACCAGTCAAGGGCAAGCTGGAGTTAATAATCTATATTTATTTGACTCAGCACT 1800  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1741 TCTCATGACCAGTCAAGGGCAAGCTGGAGTTAATAATCTATATTTATTTGACTCAGCACT 1800  
 Qy 1801 GTTTTTCATCACAACCTTGTTTTCCCGAGCATCATGTAGTGCATTTAGTTTTGTCTTTCTCAG 1860  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1801 GTTTTTCATCACAACCTTGTTTTCCCGAGCATCATGTAGTGCATTTAGTTTTGTCTTTCTCAG 1860  
 Qy 1861 GGTATAGTCAATATGCCTGCAGGAGTTTCTATAGCGAGACATAGAATAGTATTCTGATCA 1920  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1861 GGTATAGTCAATATGCCTGCAGGAGTTTCTATAGCGAGACATAGAATAGTATTCTGATCA 1920  
 Qy 1921 GTTGCCAAAGAATCTAGGAAATTAGTTGTATTTTGTGCAAGCTAATTTAAAAACATGATG 1980  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1921 GTTGCCAAAGAATCTAGGAAATTAGTTGTATTTTGTGCAAGCTAATTTAAAAACATGATG 1980  
 Qy 1981 GGCTGTTTTAAGACCAGAGTGGAATTCATGAGAGGAACTATACTACCAAAGAGCCCAA 2040  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1981 GGCTGTTTTAAGACCAGAGTGGAATTCATGAGAGGAACTATACTACCAAAGAGCCCAA 2040  
 Qy 2041 ATGACCAAATCCATGGATAATTGCTTCACAGCCTTGGCCATCCTGGCTCAGCTCTCAATT 2100  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2041 ATGACCAAATCCATGGATAATTGCTTCACAGCCTTGGCCATCCTGGCTCAGCTCTCAATT 2100  
 Qy 2101 TAGTATAATATGCAGTTCCTGTGCCTCCAGACTATGCAGCTCATCACCTAGGTTCTACA 2160  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2101 TAGTATAATATGCAGTTCCTGTGCCTCCAGACTATGCAGCTCATCACCTAGGTTCTACA 2160  
 Qy 2161 GGAAATACAGAGATGAACAACCTTGCCTTCAAAAAATGTGCTGCCTAGAAAACAGACCTG 2220  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2161 GGAAATACAGAGATGAACAACCTTGCCTTCAAAAAATGTGCTGCCTAGAAAACAGACCTG 2220  
 Qy 2221 CATTTCAACCCAACCTGTAATGCAGGATTTGGACCATGAATGATATGCTAGAATAGAAGAA 2280  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2221 CATTTCAACCCAACCTGTAATGCAGGATTTGGACCATGAATGATATGCTAGAATAGAAGAA 2280  
 Qy 2281 AGAGAAGTGTTTTTTTAATTGAGAGCCTCTATGTGCAAGGTGATATATAATCATATCCAG 2340  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2281 AGAGAAGTGTTTTTTTAATTGAGAGCCTCTATGTGCAAGGTGATATATAATCATATCCAG 2340  
 Qy 2341 TTTAATCTTCACAATATCCAATGAAGAAGGTCTCATTATCTCCATGATAAAGATGGGGAA 2400  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2341 TTTAATCTTCACAATATCCAATGAAGAAGGTCTCATTATCTCCATGATAAAGATGGGGAA 2400  
 Qy 2401 ACTAAGGTCAGAAGGGTTAACTCAACTGTCTATTGTCACATGATGAATAAATAGATGAAG 2460  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2401 ACTAAGGTCAGAAGGGTTAACTCAACTGTCTATTGTCACATGATGAATAAATAGATGAAG 2460  
 Qy 2461 TGAGATACAAAGCTGGGTTTGATTCAAAGCCCTTACTTTTCCTAATTAACTATGATGCGT 2520  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 2461 TGAGATACAAAGCTGGGTTTGATTCAAAGCCCTTACTTTTCCTAATTAACTATGATGCGT 2520

Qy 2521 ATTTATTTTTCTGCACCTTCCTTTCTTCCACAAACACATATTGATAGATGCAAGAGACTC 2580  
 |||  
 Db 2521 ATTTATTTTTCTGCACCTTCCTTTCTTCCACAAACACATATTGATAGATGCAAGAGACTC 2580  
 |||

Qy 2581 TTATTTATAAGGCGTGGGGGACAAGAAGGATACAAGGTAAGTTTCAGTGGAGCTCAGAGG 2640  
 |||  
 Db 2581 TTATTTATAAGGCGTGGGGGACAAGAAGGATACAAGGTAAGTTTCAGTGGAGCTCAGAGG 2640  
 |||

Qy 2641 ACGGGGAGATAGAACTGTGGCACTTAGGGGAGATGACATTTGCTTTGGGCAGAGGCAGCT 2700  
 |||  
 Db 2641 ACGGGGAGATAGAACTGTGGCACTTAGGGGAGATGACATTTGCTTTGGGCAGAGGCAGCT 2700  
 |||

Qy 2701 AGCCAGGACACATTTCCACTATAATTTTACAAAGTTAAATTTATAAGCTAGCATTAAAGTA 2760  
 |||  
 Db 2701 AGCCAGGACACATTTCCACTATAATTTTACAAAGTTAAATTTATAAGCTAGCATTAAAGTA 2760  
 |||

Qy 2761 AAGTGAAGTCCAGCTCCCTTGCTAAAAATAACTAGAGGTAATAATTGGTATTCAGGTAAC 2820  
 |||  
 Db 2761 AAGTGAAGTCCAGCTCCCTTGCTAAAAATAACTAGAGGTAATAATTGGTATTCAGGTAAC 2820  
 |||

Qy 2821 TCATTTACAGTCATAATGTGTTGTGAAAATTTAATCTTAAAAATTAAATTTTTAAACTAT 2880  
 |||  
 Db 2821 TCATTTACAGTCATAATGTGTTGTGAAAATTTAATCTTAAAAATTAAATTTTTAAACTAT 2880  
 |||

Qy 2881 GTGGGTCTGTGAATTTCTTTAATGTCTAAGAAATCCAGCTTCATAATTTCCATGATACAA 2940  
 |||  
 Db 2881 GTGGGTCTGTGAATTTCTTTAATGTCTAAGAAATCCAGCTTCATAATTTCCATGATACAA 2940  
 |||

Qy 2941 AGATCTTTTTTTCAGGTGGATTTTTACCTTTGTTTCCTTTTGCTCTGATAGACAAAATCAGT 3000  
 |||  
 Db 2941 AGATCTTTTTTTCAGGTGGATTTTTACCTTTGTTTCCTTTTGCTCTGATAGACAAAATCAGT 3000  
 |||

Qy 3001 TTAGGACTATTAAAGAATGTTTTGGAATAAACTGTCTTTTTCCTCAATGAATGGGATGTC 3060  
 |||  
 Db 3001 TTAGGACTATTAAAGAATGTTTTGGAATAAACTGTCTTTTTCCTCAATGAATGGGATGTC 3060  
 |||

Qy 3061 TAATGTATTTCAAATCACCCAAAACCTTTTGGCAAATAAAAAGCATTTAAAAAGAAAAAAA 3120  
 |||  
 Db 3061 TAATGTATTTCAAATCACCCAAAACCTTTTGGCAAATAAAAAGCATTTAAAAAGAAAAAAA 3120  
 |||

Qy 3121 AAAAAAAA 3128  
 |||  
 Db 3121 AAAAAAAA 3128  
 |||

SID1

SID3 - 99.0 %  
54-738-SID3  
1-685 record  
7 mismatches

[illegible]

Db	61		GCAATGGACCTCATCATGACCGTTGCTGTACTTACAATGAAAACAACCTGGTGGATGGTG	120
Qy	362		TGTATTGTCTCCCAATAGGACACTGGATTGAGGCCACTGGGCACACCAATGAAATGAAGC	421
Db	121		TGTATTGTCTCCCAATAGGACACTGGATTGAGGCCACTGGGCACACCAATGAAATGAAGC	180
Qy	422		ACACAACAGACTTCTATTTTAATATTGCAGGCCACCAAGCCATGCATTATTCAAGAATTC	481
Db	181		ACACAACAGACTTCTATTTTAATATTGCAGGCCACCAAGCCATGCATTATTCAAGAATTC	240
Qy	482		TACCAAATATCTGGCTGGGTAGCTGCCCTCGTCAGGTGGAACATGTTACCATCAAACCTGA	541
Db	241		TACCAAATATCTGGCTGGGTAGCTGCCCTCGTCAGGTGGAACATGTTACCATCAAACCTGA	300
Qy	542		AGCATGAATTGGGGATTACAGCTGTAATGAATTTCCAGACTGAATGGGATATTGTACAGA	601
Db	301		AGCATGAATTGGGGATTACAGCTGTAATGAATTTCCAGACTGAATGGGATATTGTACAGA	360
Qy	602		ATTCCTCAGGCTGTAACCGCTACCCAGAGCCCATGACTCCAGACACTATGATTAAACTAT	661
Db	361		ATTCCTCAGGCTGTAACCGCTACCCAGAGCCCATGACTCCAGACACTATGATTAAACTAT	420
Qy	662		ATAGGGAAGAAGGCTTGGCCTACATCTGGATGCCAACACCAGATATGAGCACCGAAGGCC	721
Db	421		ATAGGGAAGAAGGCTTGGCCTACATCTGGATGCCAACACCAGATATGAGCACCGAAGGCC	480
Qy	722		GAGTACAGATGCTGCCCCAGGCGGTGTGCCTGCTGCATGCGCTGCTGGAGAAGGGACACA	781
Db	481		GAGTACAGATGCTGCCCCAGGCGGTGTGCCTGCTGCATGCGCTGCTGGAGAAGGGACACA	540
Qy	782		TCGTGTACGTGCACTGCAACGCTGGGGTGGGCCGCTCCACCGCGGCTGTCTGCGGCTGGC	841
Db	541		TCGTGTACGTGCACTGCAACGCTGGGGTGGGCCGCTCCACCGCGGCTGTCTGCGGCTGGC	600
Qy	842		TCCAGTATGTGATGGGCTGGAATCTGAGGAAGGTGCAGTATTTCCCTCATGGCCAAGAGGC	901
Db	601		TCCAGTATGTGATGGGCTGGAATCTGAGGAAGGTGCAGTATTTCCCTCATGGCCAAGAGGC	660
Qy	902		CGGCTGTCTACATTGACGAAGAGGC	926
Db	661		CGGCTGTCTACATTGACGAAGAGGC	685

SID3-21

QY	54	TGGACACGTTCTGGTACAAGTTCTCTGAAGCGGGAGCCGGGAGGAGAGCTCTCTCTGGGAAG	113
Db	1	TGGACACGTTCTGGTACAAGTTCTCTGAAGCGGGAGCCGGGAGGAGAGCTCTCTCTGGGAAG	60
QY	114	GCAATGGACCTCATCATGACCGTTGCTGTACTTACAATGAAAACAACCTTGGTGGATGGTG	173
Db	61	GCAATGGACCTCATCATGACCGTTGCTGTACTTACAATGAAAACAACCTTGGTGGATGGTG	120
QY	174	TGTATTGTCTCCCAATAGGACACTGGATTGAGGCCACTGGACACACCAATGAAATGAAGC	233
Db	121	TGTATTGTCTCCCAATAGGACACTGGATTGAGGCCACTGGGCACACCAATGAAATGAAGC	180
QY	234	ACACAACAGACTTCTATTTTAATATTGCAGGCCACCAAGCCATGCATTATTCAAGAATTC	293
Db	181	ACACAACAGACTTCTATTTTAATATTGCAGGCCACCAAGCCATGCATTATTCAAGAATTC	240
QY	294	TACCAAATATCTGGCTGGGTAGCTGCCCTCGACAGGTGGAACATGTTACCATCAAACCTGA	353
Db	241	TACCAAATATCTGGCTGGGTAGCTGCCCTCGTCAGGTGGAACATGTTACCATCAAACCTGA	300
QY	354	AGCATGAATTGGGGATTACAGCTGTCTATGAATTTCCAGACTGAATGGGATATTGTTTACA	413
Db	301	AGCATGAATTGGGGATTACAGCTGTCTATGAATTTCCAGACTGAATGGGATATTGTTTACA	360
QY	414	ATTCTTCATGCTGTAACCGCTACCCAGAGCCCATGACTCCAGACACTATGATTAAACTAT	473
Db	361	ATTCTTCAGGCTGTAACCGCTACCCAGAGCCCATGACTCCAGACACTATGATTAAACTAT	420
QY	474	CTAGGGAAGAAGGCTTGGCCTACATCTGGATGCCAACACCAGATATGAGCACCCGAGGCC	533
Db	421	ATAGGGAAGAAGGCTTGGCCTACATCTGGATGCCAACACCAGATATGAGCACCCGAGGCC	480
QY	534	GAGTACAGATGCTGCCCCAGGCGGTGTGCCTGCTGCATGCGCTGCTGGAGAAGGGACACA	593
Db	481	GAGTACAGATGCTGCCCCAGGCGGTGTGCCTGCTGCATGCGCTGCTGGAGAAGGGACACA	540
QY	594	TCGTGTACGTGCACTGCAACGCTGGGGTGGGCGGCTCCACCGCGGCTGTCTGCGGCTGGC	653
Db	541	TCGTGTACGTGCACTGCAACGCTGGGGTGGGCGGCTCCACCGCGGCTGTCTGCGGCTGGC	600
QY	654	TCCAGTATGTGATGGGCTGGAATCTGAGGAAGGTGCAGTATTTCTCATGGCCAAGAGGC	713
Db	601	TCCAGTATGTGATGGGCTGGAATCTGAGGAAGGTGCAGTATTTCTCATGGCCAAGAGGC	660
QY	714	CGGCTGTCTACATTGACGAAGAGGC	738
Db	661	CGGCTGTCTACATTGACGAAGAGGC	685



5105

QY 77 TGTTACCATCAAAGTGAAGCATGAATTGGGGATTACAGCTGTCATGAATTTCCAGACTGA 136

Db	284		TGTTACCATCAA	ACTGAAGCATGA	ATTGGGGATTAC	AGCTGTAATGA	ATTTCCAGACTGA	343
Qy	137		ATGGGATATTG	TTCAGAATTC	CCTCATGCTGT	AACCGCTACCC	CAGAGCCCATG	ACTCCAGA 196
Db	344		ATGGGATATTG	TACAGAATTC	CCTCAGGCTGT	AACCGCTACCC	CAGAGCCCATG	ACTCCAGA 403
Qy	197		CACTATGATTAA	ACTATCTAGG	GAAGAAGGCTT	GGCCTACATCT	GGATGCCAAC	ACCAGA 256
Db	404		CACTATGATTAA	ACTATATAGG	GAAGAAGGCTT	GGCCTACATCT	GGATGCCAAC	ACCAGA 463
Qy	257		TATGAGCACCG	CAGGCCGAGT	ACAGATGCTG	CCCCAGGCGGT	GTGCTGCTGC	ATGCGCT 316
Db	464		TATGAGCACCG	AAGGCCGAGT	ACAGATGCTG	CCCCAGGCGGT	GTGCTGCTGC	ATGCGCT 523
Qy	317		GCTGGAGAAGG	GACACATCGT	GTACGTGCACT	GCAACGCTGGG	GTGGCCGCTCC	ACCCGC 376
Db	524		GCTGGAGAAGG	GACACATCGT	GTACGTGCACT	GCAACGCTGGG	GTGGCCGCTCC	ACCCGC 583
Qy	377		GGCTGTCTGCG	GCTGGCTCCAG	TATGTGATGGG	CTGGAATCTGAG	GAAGGTGCAGT	TATTT 436
Db	584		GGCTGTCTGCG	GCTGGCTCCAG	TATGTGATGGG	CTGGAATCTGAG	GAAGGTGCAGT	TATTT 643
Qy	437		CCTCATGGCCA	AAGAGGCCGG	CTGTCTACATT	GACGAAGAGGC	CAGCTAGCCAG	GACACATT 496
Db	644		CCTCATGGCCA	AAGAGGCCGG	CTGTCTACATT	GACGAAGAGGC	CAGCTAGCCAG	GACACATT 703
Qy	497		TCCACTATAAT	TTTTACAAAG	TAAATTTATA	AAGCTAGCATTA	AGTAAAGTGA	AGTCCAGC 556
Db	704		TCCACTATAAT	TTTTACAAAG	TAAATTTATA	AAGCTAGCATTA	AGTAAAGTGA	AGTCCAGC 763
Qy	557		TCCCTTGCTAA	AAAAATAACT	AGAGGTAATA	AATTGGTATTC	AGGTAAC	TATTTACAGTCAT 616
Db	764		TCCCTTGCTAA	AAAAATAACT	AGAGGTAATA	AATTGGTATTC	AGGTAAC	TATTTACAGTCAT 823
Qy	617		AATGTGTTGTG	AAAATTTAAT	CTTAAAAATTA	AATTTTTTAA	ACTATGTGGG	TCTGTGAAT 676
Db	824		AATGTGTTGTG	AAAATTTAAT	CTTAAAAATTA	AATTTTTTAA	ACTATGTGGG	TCTGTGAAT 883
Qy	677		TTCTTTAATGT	CTAAGAAATC	CAGCTTCATA	ATTTCCATGATA	CAAAAGATCT	TTTTTTCAG 736
Db	884		TTCTTTAATGT	CTAAGAAATC	CAGCTTCATA	ATTTCCATGATA	CAAAAGATCT	TTTTTTCAG 943
Qy	737		GTGGATTTTTA	CCTTTGTTCC	TTTTGCTCTG	ATAGACAAAAT	CAGTTTAGG	ACTATTAAA 796
Db	944		GTGGATTTTTA	CCTTTGTTCC	TTTTGCTCTG	ATAGACAAAAT	CAGTTTAGG	ACTATTAAA 1003
Qy	797		GAATGTTTTG	GGAATAAACT	GTCTTTTTTC	CTCAAT		830
Db	1004		GAATGTTTTG	GGAATAAACT	GTCTTTTTTC	CTCAAT		1037